

Available online at www.sciencedirect.com**ScienceDirect**

Procedia Computer Science 70 (2015) 325 – 333

Procedia
Computer Science4th International Conference on Eco-friendly Computing and Communication Systems

Anomaly detection in medical wireless sensor networks using machine learning algorithms

Girik Pachauri^{a*}, Sandeep Sharma^b^{a,b}*School of Information and Communication Technology
Gautam Buddha University, Greater Noida, India*

Abstract

Wireless sensor networks suffer from a wide range of faults and anomalies which hinder their smooth working. These faults are even more significant for medical wireless sensor networks, which simply cannot afford such inconsistencies. To combat this issue, various fault detection mechanisms have been developed. We tried enhancing the performance of one such mechanism, and our findings are presented in this paper. Using machine learning algorithms, we will show through our experiments on real medical datasets that our approach gives more accurate results than other existing fault detection mechanisms. This research will be critical in detecting sensor faults quickly, accurately and with a low false alarm ratio.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the Organizing Committee of ICECCS 2015

Keywords: Wireless Sensor Networks; Machine Learning Algorithms; Sensor Faults; Healthcare and patient monitoring

1. Introduction

Human longevity is steadily increasing owing to the advances of modern medicine and availability of various healthcare technologies. Technological advances coupled with the collective knowledge about human physiology

* Corresponding author.

E-mail address: girikp@gmail.com

have helped not only patients get better treatments and recover from deadly illnesses, but have also allowed doctors to make better, life-saving diagnoses in a timely manner.

However, the dearth of qualified healthcare professionals is still an impediment to the wide availability of good standard health assessment. One technological solution for this issue is wireless sensor networks. These sensor networks allow remote monitoring of patients and their real-time health stats to be readily available to the supervising physician.

Personal Area Networks (PAN) and Wireless Body Area Networks (WBAN) - the two major categories of wireless sensor network implementations in the medical sector –both consist of small wireless monitoring devices placed on the body to collect vital stats and patient metrics such as Heart Rate (HR), Blood Pressure (BP), pulse oxygen saturation (SpO₂), etc. These PANs and WBANs, however, also suffer from various issues such as faulty measurements, hardware failures, and various security issues³. While inherently limited in computational power and energy resources, measurements are further prone to a variety of anomalies including abnormal values resulting from erroneous calibration, electromagnetic interference, patients with sweating, etc., all of which may occur entirely naturally^{1,2}.

Faulty measurements degrade system accuracy and may effectuate wrong diagnoses, which may subsequently be harmful for the patients' life. Therefore, it is paramount that faulty readings be quickly and accurately detected and that they be distinguished from actual emergency situations so as to reduce false alarms.

In this paper, we will be using different machine learning algorithms to detect anomalous readings in medical WSNs. We will compare the performance of different machine learning algorithms used in our experiment to the ones used in existing techniques. We will first classify a record as normal or abnormal, then we will use regression algorithms to pinpoint the abnormal measurement in the abnormal record. We will be working on the assumption that physiological metrics are highly correlated and hence whenever genuine changes occur, they occur in two or more parameters.

The rest of this paper is organized as follows: In section II, we review related work on anomaly detection and machine learning algorithms used in medical WSNs. Section III describes briefly the machine learning algorithms used in our detection system. The mechanism's working is explained in section IV. In section V, we present the findings from our experiments conducted upon real medical dataset. Finally, we conclude the paper in section VI.

2. Related work

WSNs are becoming a major center of interest in the fields of medicine and healthcare. Various vital sign monitoring systems have been proposed, developed and deployed, such as MEDiSN⁶ & CodeBlue⁷ for monitoring HR, ECG, SpO₂ and pulse, LifeGuard⁸ for ECG, respiration & BP, Vital Jacket¹¹ for ECG & HR, etc.

A survey of security issues in healthcare applications using WSNs has been provided by P. Kumar & H. J. Lee¹². The healthcare applications of WSNs, technical challenges, and types of medical WSN systems have been reviewed and summarized by Ko et al³. R. Jurdak et al discuss anomaly based systems where they categorize the types of anomalies into Network anomalies, Node anomalies and Data anomalies¹³. Medical applications of sensor networks have also been presented in available surveys^{14,15}.

Y. Zhang et al propose a cluster based approach for detecting outliers in compromised nodes by exploiting spatiotemporal correlation and consistency¹⁶. A simple online anomaly detection algorithm based on the detection of deviation between reference and the measured time series has been proposed by Y. Yao et al¹⁷.

Interestingly, data mining techniques and machine learning algorithms have also been used in WSNs to detect anomalies in multidimensional data. For example, Naive Bayes¹⁸, Bayesian network¹⁹, Support Vector Machine (SVM)²⁰, Self-Organizing Map (SOM)²¹ which is based on neural networks, have all been used in existing anomaly detection mechanisms. In another approach, the authors used Gaussian mixture decomposition and Ant Colony algorithm to derive classification rules, which are then used to detect abnormal values²³.

In one framework each sensor applies the non-seasonal Holt-Winter algorithm to detect any deviation in the time series associated with its measurements²². Another framework uses distributed principal component analysis (DPCA) and fixed-width clustering (FWC) in order to establish a global normal profile and to detect anomalies²⁴.

One approach uses a multi-scale principal component analysis (MSPCA) based data fault detection method for WSNs²⁵. MSPCA integrates wavelet analysis and principal component analysis. A new method of anomaly detection in WSNs which classifies data using S Transform (used for feature extraction) and SVM has also been proposed²⁶.

In this paper we will apply classification algorithms such as Random Forests and k-NN and compare their performance to existing fault detection mechanisms which make use of J48. Similarly, we will apply different ensemble learning algorithms for the regression part of this framework, and compare their performance to techniques used by existing mechanisms.

3. Background

We consider the following scenario –sensing devices are attached to the patient's body for monitoring, and transmit the sensed physiological parameters to the network sink, which may be a base station or a smartphone. This base station, having higher memory, more computational capability and energy resources at its disposal, may then perform data analysis on the collected measurements to detect for anomalies or raise alarms whenever patient enters critical state, or it may store the data for later use.

The collected measurements are represented in the form of a data matrix X_{ij} , where i refers to the time instance of measurement, while j refers to the measured parameter. Equation 1 shows the data matrix structure:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad (1)$$

We will use a classification algorithm to detect abnormal records. A record refers to the collection of measurements of different parameters at the same instant, i.e., a row in the data matrix X . We will then use a regression algorithm to measure the correlation between predicted and actual values of the parameters. This is done to see if the difference between predicted and actual values does not exceed the threshold, which we have assumed to be 10% here. If a reading exceeds this threshold, then correlation analysis is done to differentiate between a faulty reading and patient entering into critical state. In the rest of this section, we will discuss the algorithms used and their working in brief.

3.1. J48 Decision Trees

A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute (monitored physiological attribute, here), each branch represents the outcome of the test and each leaf node represents a class label (normal or abnormal). Classifying test data is straightforward once a decision tree has been constructed. The criteria used to select the best attribute for splitting the data is Gain Ratio, which can be calculated as:

$$GR(X, X_k) = \frac{IG(X, X_k)}{SI(X, X_k)} \quad (2)$$

Here, SI stands for the split information. This information is sensitive to how broadly and uniformly the attribute splits the data. IG stands for information gain, and can be calculated as:

$$IG(X, X_k) = H(X) - \sum_{x_{ik} \in X} \frac{|x_{ik}|}{|X|} H(x_{ik}) \quad (3)$$

where \mathbf{X} is the dataset, \mathbf{X}_k is a column in the dataset, $H(X)$ is the entropy of the dataset, and x_{ik} are the values taken by attribute \mathbf{X}_k .

Similarly, SI can be calculated as:

$$SI(X, x_{ik}) = - \sum_{c=1}^n \frac{|x_{ik}|}{|X|} \log_2 \frac{|x_{ik}|}{|X|} \quad (4)$$

where n is the number of classes, and $SI(X, x_{ik})$ is the entropy of the instance x_{ik} within each class. Once we know the gain ratio for each attribute, we will be able to hierarchically distribute those attributes into the tree nodes.

3.2. Random Forests

The random forest²⁷ is an ensemble approach primarily based on decision tree which, in ensemble terms, corresponds to a weak learner. Random forests help by averaging multiple deep decision trees, trained on the same training set but on different parts, with the aim of reducing the variance so that there is no overfitting of training sets.

The process flow is as given: A new input fed into the system is run through every tree. We may get the result as either an average or a weighted average of all of the leaf nodes that are reached, or, in the case of categorical variables, a voting majority. However, greater the inter-tree correlation, greater the random forest error rate. Therefore, the model would be well-suited to have the trees as uncorrelated as possible.

- Advantage: Random forest runtimes are quite fast, and they can even handle unbalanced or missing data.
- Disadvantage: A weakness of random forests is that when used for regression, they cannot predict beyond the range of the training data. They may also over-fit data sets that are particularly noisy.

3.3. k-Nearest Neighbours

Instance-based classifiers such as kNN operate on the principle that unknown instances can be classified by relating the unknown to the known according to some distance/similarity function. The logic here is that instances far apart in the instance space defined by the distance function are less likely than closely situated instances to belong to the same class.

Classification using an instance-based classifier is just a matter of locating the nearest neighbour in instance space and labelling the unknown instance with the same class label as that of the located (known) neighbour. This approach is often referred to as a nearest neighbour classifier. Nearest neighbour classifiers are highly susceptible to noise in the training data due to the high degree of local sensitivity.

3.4. Linear Regression

Linear regression is an approach for modeling the relationship between a scalar dependent variable y and one or more independent variables denoted by \mathbf{X} . This statistical method models the dependent variable y_{ik} using a vector of independent variables x_{ik} called regressors. The model is represented as :

$$y_{ik} = C_0 + C_1 x_{i1} + C_2 x_{i2} + \dots + C_n x_{in} \quad (5)$$

where y_{ik} is the dependent variable at instance i , x_{ik} are the regressors and C_n are the coefficients of the regressors (weights). These coefficients are calculated in the training phase as the covariance of \mathbf{X}_k and \mathbf{Y}_k is divided by the variance \mathbf{X}_k .

$$C_k = \frac{Cov(X_k, Y_k)}{Var(X_k)} = \frac{\sum (x_{ik} - \bar{X})(y_{ik} - \bar{Y})}{\sum (x_{ik} - \bar{X})^2} \quad (6)$$

This process is done to predict the value of y_{ik} by using other attributes in the same instance $x_{ij|j \neq k}$, and to compare the predicted y_{ik} with the actual value of x_{ik} to find if it fits within the expected margin of error.

3.5. Additive Regression

This method represents a generalization of multiple regression (which is a special case of general linear models). If we maintain the additive nature of the model, while replacing the simple terms of the linear equation $\sum C_i x_i$ with $f_i(x_i)$ where f_i is a non-parametric function of the predictor x_i , that would be a generalization of the multiple regression model. Therefore it can be said that in additive models an unspecified non-parametric function is estimated for each predictor, to achieve the best prediction of the dependent variable values, while in linear models, a single coefficient for each variable (additive term) is estimated.

Broadly speaking, the term “additive regression” refers to any way of generating predictions by summing up contributions obtained from other models. Most learning algorithms for additive models do not build the base models independently but ensure that they complement one another and try to form an ensemble of base models such that the predictive performance is optimized according to some specified criterion.

3.6. Decision Stump

The decision stump is a machine learning technique which basically consists of a decision tree with one internal node (the root) which is immediately connected to the terminal nodes (its leaves). Prediction is made based on the value of just a single input feature. Sometimes they are also called 1-rules.

Several variations are possible depending on the type of the input feature. For nominal features, a stump may be built containing a leaf for each possible feature value or a two-leave stump, with one leave corresponding to some chosen category, and the other leaf to all the other categories. These two schemes are identical for binary features.

Usually, for continuous features, some threshold feature value is selected, and the stump contains two leaves: for values below and above the threshold. However, rarely, multiple thresholds may be chosen and the stump therefore may contain three or more leaves. Machine learning ensemble techniques such as bagging and boosting often employ decision stumps components (called “weak” or “base” learners).

4. Implementation

The assumed scenario has already been discussed previously. The detection mechanism has two parts: a classification problem, and a regression problem. First, it will classify a record as normal or abnormal, and then for each abnormal record it will pinpoint exactly which parameter crosses the threshold. This will be done so that we can further perform correlation analysis to distinguish between faulty readings and patient entering critical state. In the rest of the paper, we will be focusing on only the following attributes: heart rate (HR), pulse oxygen saturation (SpO2), PULSE, body temperature (Tblood), and respiration rate (RESP).

The mechanism will be divided into 2 working phases: first a model will be built to classify the data. Then, we will feed the records into the model as inputs to be classified as normal or abnormal. Classification will be performed by the following classification algorithms: J48, Random Forests, and k-Nearest Neighbours. Their performances will be compared to determine which performed better.

Once an abnormal record has been identified, we will then apply regression algorithms such as Linear Regression and Additive Regression so as to further perform correlation analysis and distinguish between faulty measurements and actual critical state. Results from these regression algorithms will be compared to see which one performs better.

Classification is done on the dataset to greatly reduce the complexity, so that regression need not be applied on every attribute for each instance. The reader is advised to refer to *Data Mining: Practical Machine Learning Tools and Techniques*²⁸ to get a more detailed explanation of all the algorithms used.

5. Experimental results

The dataset used in our research has been obtained from PhysioNet⁴, an online database of recorded physiological signals. We will be using the MIMIC dataset, which contains 121 records and each recording contains total of 12 attributes: ABPmean, ABPsys, ABPdias, C.O., HR, PAPmean, PAPsys, PAPdias, PULSE, RESP, SpO2, Tblood.

To measure the performance and test the efficiency of the different algorithms, we compare the results through the use of the WEKA⁵ tool. After applying the different classification algorithms, the results obtained are as follows:

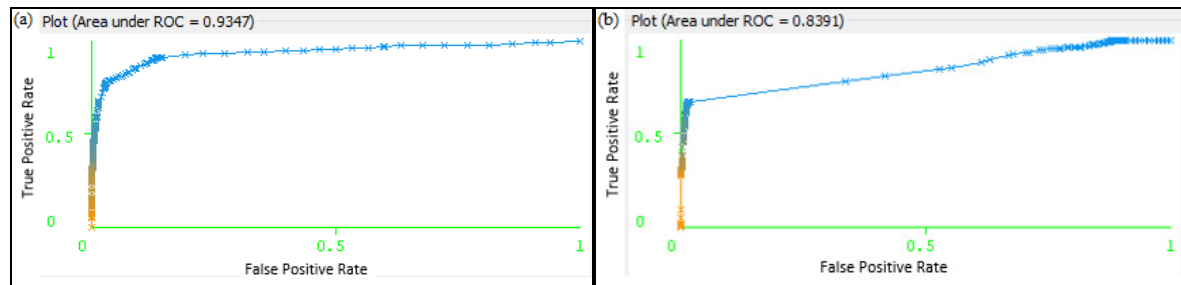


Fig. 1 (a) J48 ROC Curve; (b) k-NN ROC Curve

Fig.1. (a) and (b) show the Receiver Operating Characteristics (ROC) curve of J48 and k-NN algorithms. The ROC curve illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve plots the true positive rate (also called sensitivity) against the false positive rate (also called specificity) at various thresholds. The ROC curve is thus the sensitivity as a function of specificity.

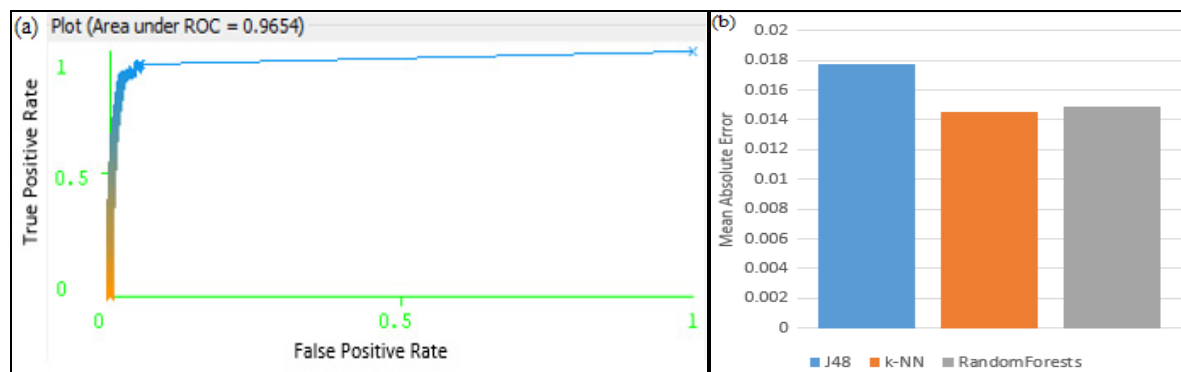


Fig. 2 (a) Random Forests ROC Curve; (b) Mean Absolute Error of Different Classifiers

Fig.2. (a) shows the ROC curve of Random Forests algorithm. It can be clearly seen that the curve for Random Forests is the best among the three. The area under ROC, which shows the overall performance of a classifier is also evidently largest for Random Forests algorithm.

Fig.2 (b) shows the mean absolute error for each classifier. Once again, Random Forests offers good performance, along with k-NN algorithm, though that is true only for small datasets like these. The mean error for both k-NN and Random Forests algorithms is much less than J48 even though k-NN misclassifies a lot more instances than J48.

After applying these classification algorithms, we now approach the regression part of the system. We applied different regression algorithms, and the results are shown in Fig.3.

Of all the applied algorithms, the least mean error is provided by the meta-learner scheme of Additive Regression while using k-NN as base learner. Furthermore, this same scheme also provides the best correlation coefficient among all the applied algorithms.

Finally, a comparison of the run-times of the various algorithms has been presented in Tables 1 & 2. Table 1 shows that for classification, k-NN takes the least time to build the model, followed by J48, and finally Random Forests. However, it is necessary to remember that k-NN had the most misclassifications. Random Forests takes more time than J48 even though both of them are based on the decision tree approach, because unlike J48, it gives final output as a weighted average of all the leaf nodes.

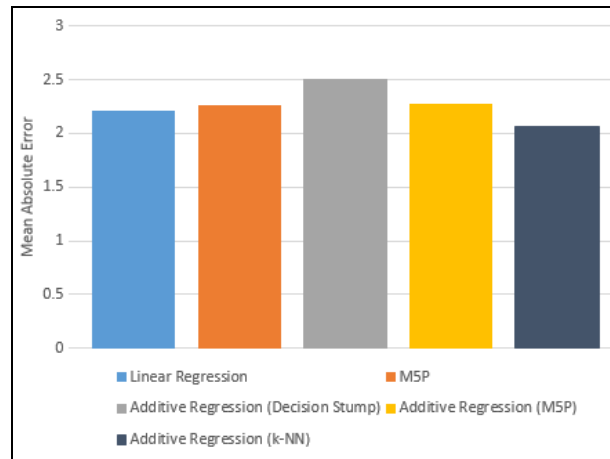


Fig. 3. Mean Absolute Error for Different Regression Algorithms

Similarly, Table 2 shows the run-times of regression algorithms. Here, Linear Regression takes the least time, while Additive Regression (with k-NN) takes the most time to build the model. However, once again, it is important to keep in mind that Additive Regression (with k-NN) had the least mean error.

Table 1. Classification algorithm run-times.

Algorithm	Time (in seconds)
J48 Decision Tree	0.19
k-Nearest Neighbours	0.02
Random Forests	1.43

Table 2. Regression algorithm run-times.

Algorithm	Time (in seconds)
Linear Regression	0.22
M5P	0.73
Additive Regression (Decision Stump)	0.39
Additive Regression (M5P)	2.04
Additive Regression (k-NN)	4.87

Summarizing the results, we can say that taking into account both the error performances and run-time performances of all the algorithms, Random Forests provides the best overall performance for classification, while Additive Regression (with k-NN) performs best for regression jobs.

6. Conclusions

From the experiments performed, we can safely say that machine learning techniques and algorithms can play a significant part in devising a fault and anomaly detection framework for use in medical wireless sensor networks.

The proposed framework integrates Random Forests algorithm for classification jobs and Additive Regression techniques for prediction jobs for anomaly detection in medical WSNs. This approach achieves both spatial and temporal analysis for anomaly detection. We have tested this framework on real medical dataset available from reliable sources, and it has been shown that both these algorithms perform much better than other previous research techniques.

With the growing computational capabilities of modern computers, and the equally rapid adaptation of WSNs in different walks of life such as medicine and health care, we can only surmise that the use of machine learning in the medical field is only about to expand and make its presence felt even more.

References:

1. Chipara, C. Lu, T. C. Bailey, and G. C. Roman, "Reliable Clinical Monitoring using Wireless Sensor Networks: Experiences in a Step-down Hospital Unit," in Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems (SenSys'10), pp. 155–168, 2010.
2. Y. Zhang, N. Meratnia, and P. J. M. Havinga, "Outlier Detection Techniques for Wireless Sensor Networks: A Survey", IEEE Communications Surveys and Tutorials, vol. 12, no. 2, pp. 159–170, 2010.
3. J. Ko, C. Lu, M. B. Srivastava, J. A. Stankovic, A. Terzis, and M. Welsh, "Wireless Sensor Networks for Healthcare", Proceedings of the IEEE, vol. 98, no. 11, pp. 1947–1960, 2010.
4. "Physionet," <http://www.physionet.org/cgi-bin/atm/ATM>.
5. "Weka data mining tool," <http://www.cs.waikato.ac.nz/~ml/weka/>.
6. J. Ko, J. H. Lim, Y. Chen, R. Musvaloiu-E, A. Terzis, G. M. Masson, T. Gao, W. Destler, L. Selavo, and R.P. Dutton, "Medisn: Medical emergency detection in sensor networks," ACM Transactions on Embedded Computing Systems (TECS), vol. 10, no. 1, pp. 1–29, 2010.
7. D. Malan, T. Fulford-jones, M. Welsh, and S. Moulton, "CodeBlue: An Ad Hoc Sensor Network Infrastructure for Emergency Medical Care," in Proceedings of International Workshop on Wearable and Implantable Body Sensor Networks, 2004.
8. K. Montgomery, C. Mundt, G. Thonier, U. Udoh, V. Barker, R. Ricks, L. Giovangrandi, P. Davies, Y. Cagle, J. Swain, J. Hines, and G. Kovacs, "Lifeguard – A personal physiological monitor for extreme environments," in Proceedings of the IEEE 26th Annual International Conference on Engineering in Medicine and Biology Society, pp. 2192–2195, 2004.
9. O. Salem, A. Guerassimov, A. Mehaoua, A. Marcus, and B. Furht, "Sensor Fault and Patient Anomaly Detection and Classification in Medical Wireless Sensor Networks", IEEE ICC 2013 - Selected Areas in Communications Symposium, 2013.
10. S. J. Hua, M. C. Xiang, "Anomaly Detection Based on Data-Mining for Routing Attacks in Wireless Sensor Networks", Second International Conference on Communications and Networking in China (CHINACOM '07), 2010.
11. J. P. S. Cunha, B. Cunha, A. S. Pereira, W. Xavier, N. Ferreira, and L. Meireles, "Vital-JacketR: A wearable wireless vital signs monitor for patients' mobility in cardiology and sports," in International Conference on Pervasive Computing Technologies for Healthcare, PervasiveHealth, 2010.
12. P. Kumar and H.-J. Lee, "Security Issues in Healthcare Applications Using Wireless Medical Sensor Networks: A Survey," Sensors, vol. 12, no. 1, pp. 55–91, 2012.
13. R. Jurdak, X. R. Wang, O. Obst, and P. Valencia, Wireless Sensor Network Anomalies: Diagnosis and Detection Strategies. Springer, vol. 10, ch. 12, pp. 309–325, 2011.
14. K. Grgic, D. ˇ Zagar, and V. Križanovic, "Medical applications of wireless sensor networks – current status and future directions," Medicinski Glasnik, vol. 9, no. 1, pp. 23–31, 2012.
15. H. Alemdar and C. Ersoy, "Wireless sensor networks for healthcare: A survey," Computer Networks, vol. 54, no. 15, pp. 2688–2710, 2010.
16. Y. Zhang, H.-C. Chao, M. Chen, L. Shu, C. hyun Park, and M.-S. Park, "Outlier Detection and Countermeasure for Hierarchical Wireless Sensor Networks," IET Information Security, 2009.
17. Y. Yao, A. Sharma, L. Golubchik, and R. Govindan, "Online Anomaly Detection for Sensor Systems: a Simple and Efficient Approach," Performance Evaluation, vol. 67, no. 11, pp. 1059–1075, 2010.
18. X. Yang, A. Dinh, and L. Chen, "Implementation of a Wearable Real-Time System for Physical Activity Recognition based on Naive Bayes Classifier," in International Conference on Bioinformatics and Biomedical Technology (ICBBT'10), 2010.
19. A. Farruggia, L. R. Giuseppe, and M. Ortolani, "Probabilistic Anomaly Detection for Wireless Sensor Networks," in Proceedings of the 12th international conference on Artificial intelligence around man and beyond, pp. 438–444, 2011.
20. A. S. Raghuvanshi, R. Tripathi, and S. Tiwari, "Machine Learning Approach for Anomaly Detection in Wireless Sensor Data," International Journal of Advances in Engineering & Technology, vol. 1, no. 4, pp. 47–61, 2011.
21. S. Siripanadorn, W. Hattagam, and N. Teaumroong, "Anomaly Detection in Wireless Sensor Networks using Self-Organizing Map and Wavelets," International Journal of Communications, vol. 4, no. 3, pp. 74–83, 2010.
22. O. Salem, Y. Liu, and A. Mehaoua, "Detection and Isolation of Faulty Measurements In Medical Wireless Sensor Networks" in First International Symposium on Future Information and Communication Technologies for Ubiquitous HealthCare (Ubi-HealthTech), 2013.
23. A. Naseem, O. Salem, Y. Liu, and A. Mehaoua, "Reliable Vital Sign Collection in Medical Wireless Sensor Networks" in IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013), 2013.
24. M. A. Livani, M. Abadi, "Distributed PeA-based Anomaly Detection in Wireless Sensor Networks", International Conference for Internet Technology and Secured Transactions (ICITST), 2010.

25. X. Y. Xin, C. X. Guang, Z. Jun,” Data Fault Detection for Wireless Sensor Networks Using Multi-scale PCA Method”, Second International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), 2011.
26. Arpita Bhargava, A.S. Raghuvanshi,” Anomaly Detection in Wireless Sensor Networks using S-Transform in combination with SVM”, 5th International Conference on Computational Intelligence and Communication Networks, 2013.
27. L. Breiman, Random Forests, Machine Learning, October 2001, Volume 45, Issue 1, pp 5-32, 2001.
28. I. H. Witten, E. Frank, and M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques (Third Edition). Morgan Kaufmann Publishers Inc., 2011.